Computer Vision and Image Understanding 144 (2016) 205–216

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Efficient multi-target tracking via discovering dense subgraphs

Behzad Bozorgtabar [a], Roland Goecke [a,b,*]

[a] Vision & Sensing, Human-Centred Technology Research Centre, University of Canberra, Australia
[b] HCC, CECS, Australian National University, Australia

ARTICLE INFO

Article history:
Received 25 December 2014
Accepted 25 November 2015

Keywords:
Multi-target tracking
Dense subgraphs
Relation affinity graph
Discriminative target appearance model

ABSTRACT

In this paper, we cast multi-target tracking as a dense subgraph discovering problem on the undirected relation graph of all given target hypotheses. We aim to extract multiple clusters (dense subgraphs), in which each cluster contains a set of hypotheses of one particular target. In the presence of occlusion or similar moving targets or when there is no reliable evidence for the target's presence, each target trajectory is expected to be fragmented into multiple *tracklets*. The proposed tracking framework can efficiently link such fragmented target trajectories to build a longer trajectory specifying the true states of the target. In particular, a discriminative scheme is devised via learning the targets' appearance models. Moreover, the smoothness characteristic of the target trajectory is utilised by suggesting a smoothness tracklet affinity model to increase the power of the proposed tracker to produce persistent target trajectories revealing different targets' moving paths. The performance of the proposed approach has been extensively evaluated on challenging public datasets and also in the context of team sports (e.g. *soccer*, *AFL*), where team players tend to exhibit quick and unpredictable movements. Systematic experimental results conducted on a large set of sequences show that the proposed approach performs better than the state-of-the-art trackers, in particular, when dealing with occlusion and fragmented target trajectory.

1. Introduction

Multi-target tracking in real world scenarios is a crucial problem for many computer vision tasks. Some of its potential applications include anomaly detection, visual surveillance, human–computer interaction and sports analysis.

Generally speaking, a multi-target tracking framework takes a set of target detections in each frame from any pre-trained detector as its input and aims to recover the trajectories of all targets as well as maintaining their consistent identities. However, the problem and its difficulty depend on conflicting challenges, such as occlusion, drastically varying illumination and viewpoint for the target during the tracking, and therefore its trajectory is expected to be fragmented into multiple tracklets across video sequence. In addition, recovering the true target trajectories is an ambiguous problem when several visually similar targets move closely together. Apart from these challenges, the detector cannot be assumed to be accurate and may produce false negatives (missed targets) or extra detections. In such scenarios, revealing each individual's identity results in an extreme complexity in the number of tracks and measurements, and thus quickly becomes infeasible.

Recently, *tracking-by-detection* methods [1,2] have attracted much attention due to the promising improvements on object detection. They usually seek proper association between the target hypotheses by constructing their mutual similarities (affinities) based on multiple cues, such as *motion similarity*.

In the proposed framework, multi-target tracking is formulated as a dense subgraph extraction problem, in which each recovered dense subgraph specifies the trajectory of the one particular target during the tracking. Here, the higher order correspondences between the observations (targets' hypotheses across video sequence) are considered. With respect to the literature, this approach has the following advantages:

1. An iterative strategy is devised to partition the spatio-temporal graph of target hypotheses into different subgraphs. Each reveals a different target trajectory. At each iteration, the most coherent vertices of the constructed spatio-temporal graph in a short time period of the video will be added to the existing subgraphs (tracklets) until reaching the densest subgraphs, which are more likely to be the true trajectories of the targets. Since the proposed method works on small graph partitions (in the few video frames), it is much more efficient in terms of computational expense.
2. In the presence of occlusion of the target, whether caused by another target or visual obstacles where the target trajectories are only partially observable due to a failure of the pre-trained

* Corresponding author at: Vision & Sensing, Human-Centred Technology Research Centre, University of Canberra, Canberra, ACT 2601, Australia.
E-mail address: roland.goecke@ieee.org (R. Goecke).

http://dx.doi.org/10.1016/j.cviu.2015.11.013
1077-3142/

Fig. 1. Some tracking results from a '*TUD-Campus*' video sequence without (*first row*) and with (*second row*) considering higher order relationships between different detections across the temporal domain. The proposed tracker (*second row*) is able to handle mutual occlusion of a walking person caused by another similar person, while the target is assigned a new ID (highlighted with different colour) in the first row obtained by locally constrained-temporal-domain methods, which associate targets in a limited time period of the sequence.

detector, a discriminative scheme is devised via learning the targets' appearance models. In addition, target hypotheses are sparse in time (*false negative*), which prevents obtaining good target motion estimates. Therefore, the smoothness characteristic of the target trajectory is utilised by suggesting a smoothness tracklet affinity model to increase the power of the proposed tracker to associate the fragmented target tracklets to produce persistent target trajectories revealing different targets' moving paths (see Fig. 1).

2. Related work

Most of the recent approaches to tracking which pursue a *tracking-by-detection* strategy, comprises two steps: (i) obtaining a set of independent target candidates in each frame (*detection*) and (ii) assigning the target identities to these candidates (*data association*). Although there are some related work [3] which address both problems (object detection and data association) via proposing these two steps in a single framework, but the majority of state-of-the-art methods considered these two problems separately and the priority is usually given to the data association design. Therefore, the related work in this tracking step is extensively described here:

2.1. Local data association

These methods consider correspondence between the target observations (detections' responses) in the temporally local frames which yields the polynomial complexity. The best example of such approaches are Bi-partite matching methods [4,5]. Brendel et al. [5] presented an approach to formulate data association problem as finding the *Maximum-Weight Independent Set* (MWIS) of the graph of detections' responses. Shu et al. [4] suggested part-based model to obtain human part detections, which is robust against partial occlusion and appearance changes. Some methods devise a greedy strategy for the data association by solving a series of Bi-partite assignment problems to assign an evolving set of target trajectories to the target hypotheses in each video frame. Shafique and Shah [6] proposed a *k-partite* complete graph where the connections in the graph are not limited to only two consecutive frames. Bae and Yoon [7] proposed an online multi-object tracking system with a new data association technique considering the track existence probability. Although this class of methods are computationally efficient, but they fail in tracking visually similar moving targets or when there exists heavy occlusion or pose variations, due to their limited-temporal-locality strategies.

2.2. Global data association

Unlike local association based methods, recently, another category of data association based techniques (namely *global association*) have been proposed which are by far, the most explored strategy for multi-target tracking. The global association approaches operate on the *batch* of frames and sometimes the whole temporal sequence at once [8,9]. These approaches formulate the higher order relationships between different target observations (detection responses).

Multi-object tracking has been recently formulated as a *network flow* problem where a set of object tracks are resolved simultaneously by solving min-cost flow techniques. Different approaches to minimum cost flow have been presented recently. Zhang et al. [10] proposed a global optimal solution for the network flow optimisation using *push-relabel* algorithm. Pirsiavash et al. [11] utilise the same graph as in [10], and use a fast greedy shortest path algorithm for the tracking problem. Berclaz et al. [12] suggested a globally tracking by relying on the *k-shortest paths* (KSP) algorithm to solve the flow problem. To handle the difficulties in object localisation caused by occlusion and clutter in the video, recently Chari et al. [13] proposed to add pairwise costs to the min-cost network flow framework where a convex relaxation strategy with an efficient *rounding heuristic* is designed to solve the problem. The main drawback of these approaches is that, they do not consider acceleration information for computing trajectories which is required for trajectories' smoothness in spatio-temporal domain.

Work by Schindler [14] and Milan et al. [15] proposed multi-target tracking in crowded scenarios by incorporating mutual exclusion between the targets at both: *trajectory estimation* where any two trajectories should remain spatially separated as well as the *data association* step, in which each trajectory should be assigned at most one detections per frame. However, for this purpose, they employed an already-performed tracking from [11] as an input to their approach. Hence, these work should rather be considered as trajectory-refinement procedure.